

Sprawozdanie z realizacji zadania badawczego nr 11 za 2023 rok

Tytuł zadania **Identyfikacja mechanizmu molekularnego odporności żyta ozimego na rdzę brunatną.**

Kierownik zadania: **dr hab. inż. Kamila Nowosad , prof. Uczelni**

Celem prac realizowanych w 2023 roku było:

- 1) Wytypowanie genotypów żyta o największej tolerancji lub całkowitej odporności na porażenie rdzą brunatną (*Puccinia recondita*).
- 2) Uzyskanie informacji o genotypie osobników populacji F₂ niezbędnych do utworzenia mapy genetycznej.
- 3) Identyfikacja genów o zmienionej ekspresji po infekcji patogenem grzybowym roślin wybranych z materiałów kolekcyjnych hodowców.

W roku 2023 wykonano następujące prace:

Ad 1)

Materiał badawczy stanowiły przekazane przez hodowców materiały wyjściowe do hodowli nowych odmian żyta ozimego, linie wsobne z kolekcji oraz zarejestrowane odmiany, które zostały użyte jako wzorce.

1. Formy żyta ozimego przekazane przez:

1.1. DANKO Hodowla Roślin Sp. z o. o. (60 genotypów)

- A) S7N/23, S12N/23, S17N/23, S37N/23, S39N/23, S63N/23, S65N/23, S80N/23, S82N/23, S83N/23, S89N/23, S98N/23, S105N/23, S109N/23, S113N/23, S123N/23, S127N/23, S129N/23, S134N/23, S135N/23
- B) SOA 1/23, SOA 2/23, SOA 3/23, SOA 4/23, SOA 5/23, SOA 6/23, SOA 7/23, SOA 8/23, SOA 9/23, SOA 10/23, SOA 11/23, SOA 12/23, SOA 13/23, SOA 14/23, SOA 15/23, SOA 16/23, SOA 17/23, SOA 18/23, SOA 19/23, SOA 20/23
- C) Szk 03/20R, Szk 04/21R, Szk 05/21R, Szk 06/21RS, Szk 07/22R, Szk 09/19R, Szk 10/20RS, Szk 11/22R, Szk 12/19R, Szk 13/21R, Szk 14/21R, Szk 15/19R, Szk 16/22R, Szk 17/19R, Szk 18/21RS, Szk 19/20K, Szk 20/22RS, Szk 21/22K, Szk 22/20K, Szk 23/21K

1.2. Hodowla Roślin Smolice Sp. z o. o. Grupa IHAR (60 genotypów)

HRSM 637 - 4R, HRSM 658 - 4R, HRSM 661 - 4R, HRSM 647 - 4R, HRSM 693 - 3R, HRSM 701 - 3R, HRSM 706 - 2R, HRSM 711 - 2R, HRSM 718 - 2R, HRSM 720 - 2R, HRSM 726 - 2R, HRSM 734 - 2R, HRSM 736 - 2R, HRSM 754 - 2R, HRSM 765 - 2R, HRSM 766 - 2R, HRSM 776 - 2R, HRSM 785 - R, HRSM 786 - R, HRSM 790 - R, HRSM 804, HRSM 805, HRSM 806, HRSM 807, HRSM 808, HRSM 809, HRSM 810, HRSM 811, HRSM 813, HRSM 814, HRSM 815, HRSM 816, HRSM 817, HRSM 818, HRSM 819, HRSM 820, HRSM 821, HRSM 822, HRSM 823, HRSM 824, HRSM 826, HRSM 827, HRSM 828, HRSM 830, HRSM 831, HRSM 832, HRSM 833, HRSM 834, HRSM 835, HRSM 836, HRSM 837, HRSM 838, HRSM 839, HRSM 840, HRSM 842, HRSM 843, HRSM 844, HRSM 845, HRSM 846, HRSM 847

1.3 Poznańska Hodowla Roślin Sp. z o.o. (60 genotypów)

NS25N/14, NS0858N/05, NS419N/1/20, NS1832N/1, NS151N/14, NS1120N/4, 857/2N/21, NS1N/22, NS2N/22, NS5N/22, NS10N/22, NS11P/22, NS13P/22, NS14N/22, NS15P/22, NS16N/22, NS17N/22, NS22N/22, NS23N/22, NS24N/22, NS25N/22, NS26N/22, NS28N/22,

NS98N/14, NS33N/16, NS33N/22, NS36N/22, 3503/3, 4374/1, 4464/1, 3527/4, 3535/1, 3545/2, 3558/1, 3636/4, 3666/1, 3697/1, 3715/3, 3734/3, 3734/4, 3805/1, 3730/1, 3833/1, 3839/1, 3856/1, 3877/1, 3882/1, 3889/3, 3911/1, 3918/1, 3970/1, 3984/1, 4000/2, 4002/1, 4014/1, Ekran 37/22, 4073/1, 4099/1, 4139/1, 4150/5

Z przesłanych materiałów do badań 10 genotypów nie kiełkowało i nie zostały one poddane analizie, były to:

NS23N/22, NS10N/22, NS1N/22, HRSM 809, HRSM 814, HRSM 637 - 4R, NS16N/22, 3527/4, 4073/1, HRSM 754 - 2R

W roku 2023 testowaniu w kierunku poszukiwania odporności na rdzę brunatną poddano łącznie 170 genotypów żyta ozimego plus wzorce w doświadczeniach infekcyjnych. Były to dostarczone przez hodowców linie wsobne, populacje oraz odmiany wzorcowe.

Ocenę podatności genotypów żyta na porażenie przez rdzę brunatną (*Puccinia recondita*) wykonano metodą laboratoryjną. Na szlakach Petriego przygotowano pożywkę składającą się z agaru z dodatkiem benzomidazolu. Liście dziesięciodniowych siewek układano na pożywce, a następnie infekowano przygotowanym wcześniej inokulatem rdzy brunatnej. Materiał do przygotowania zawiesiny inokulacyjnej zebrano w różnych miejscowościach (m.in: Wrocław, Smolice, Kondratowice, Radzików), aby otrzymać szeroką genetycznie populację zarodników patogena. Oceny porażenia wykonano po dziesięciu dniach od inokulacji, w skali czterostopniowej, gdzie 1 – brak porażenia, a 4 - silne objawy porażenia rdzą brunatną na liściu.

Wyniki ocen zostały poddane analizom statystycznym. W celu prawidłowego statystycznego opracowania otrzymanych danych eksperymentalnych przeprowadzono ich transformację, w celu spełnienia warunków koniecznych do analiz, czyli uzyskania ciągłości i normalności rozkładu empirycznego. Wyniki surowe przekształcono wykorzystując wzór podany przez Węgrzyna i in. (1996), a następnie przeprowadzono jednoczynnikową analizę wariancji zgodnie z modelem kompletnej randomizacji (ANOVA), co umożliwiło sprawdzenie istotności zróżnicowania badanych genotypów żyta ozimego pod względem odporności na rdzę brunatną. Do grupowania genotypów o tej samej wartości średnich, czyli do ich podziału na grupy jednorodne wykorzystano wielokrotny test Duncana.

W doświadczeniach infekcyjnych wykonano oceny podatności posiadanych form żyta na porażenie przez *Puccinia recondita*. Oceną objęto łącznie 170 genotypów żyta plus formy kontrolne - linie UPWR.

Wyniki ocen dla badanych genotypów żyta zostały poddane analizie statystycznej. Do analizy wariancji oraz podziału genotypów na grupy jednorodne wykorzystano dane po transformacji, które po pogrupowaniu zostały z powrotem przeliczone na skalę oceny.

Analiza wariancji ujawniła znaczące różnice w odporności badanych genotypów żyta na rdzę brunatną. Oznacza to, że stopień porażenia liści przez rdzę brunatną podczas sztucznej inokulacji różnił się w zależności od genotypu. Spośród 170 analizowanych genotypów żyta, zidentyfikowano 14 linii, które wykazały pożądaną, wysoką odporność na tego patogena. Wiele badanych genotypów było podatnych na infekcję i oceniono je na 3 lub 4 w stosowanej czterostopniowej skali oceny porażenia. Jednakże 14 linii (**S63N/23, S80N/23, S113N/23, SOA 3/23, Szk 07/22R, HRSM 658 - 4R, HRSM 701 - 3R, HRSM 720 - 2R, HRSM 785 - R, NS419N/1/20, NS151N/14, 3574/1, 4014/1, 4000/2**) wykazało minimalne porażenie lub całkowity brak zarodników na liściach, co pozwala uznać je za odporne na rdzę brunatną. Te genotypy stworzyły pierwszą jednorodną grupę. Pozostałe genotypy, które w badaniach laboratoryjnych uzyskały wyższe oceny, tworzyły kolejne grupy jednorodne, wskazujące na ich większą wrażliwość na rdzę brunatną. Zgodnie z literaturą, istnieje wysoka korelacja pomiędzy odpornością określoną w warunkach laboratoryjnych a odpornością w terenie. Wybrane genotypy mogą być wykorzystane w programach hodowli nowych odmian żyta tolerancyjnych na rdzę brunatną.

Wyniki z bieżącego roku dotyczące oceny porażenia genotypów żyta przez rdzę brunatną są zgodne z wcześniejszymi badaniami [Sawinska 2011; Pietrusińska i in 2018]. Spośród przebadanych genotypów żyta ozimego, zaobserwowano kilka form całkowicie odpornych na rdzę brunatną. Zasięg porażenia przez patogena w badanych liniach wahał się od 1,0 do 4,0. Przeprowadzona analiza statystyczna wykazała znaczące różnice w podatności na rdzę brunatną wśród badanych genotypów. Niektóre formy wykazały odporność na *Puccinia recondita*. Podobne obserwacje dotyczące linii wsobnych pochodzą z badań przeprowadzonych przez Katedrę Genetyki, Hodowli Roślin i Nasiennictwa Uniwersytetu Przyrodniczego we Wrocławiu. Również linie wsobne żyta ozimego, będące częścią całej kolekcji, wykazały istotne różnice w podatności na rdzę brunatną [Weber i in. 2015; Bujak i in. 2019].

Ad 2)

Materiał badawczy stanowiło genomowe DNA roślin pokolenia F₂, które uzyskano w poprzednim roku badań. Genomowe DNA wyizolowano za pomocą komercyjnego kitu do ekstrakcji kwasów nukleinowych, zgodnie z instrukcją producenta. Zastosowanie komercyjnego kitu zapewniło izolację wysokiej jakości DNA o zbliżonym stężeniu dla wszystkich analizowanych obiektów. Pierwotnie zaproponowany wariant przewidywał genotypowanie markerów SNP za pomocą techniki targetGBS. W 2020 roku ta technika wydawała się najbardziej korzystnym cenowo rozwiązaniem umożliwiającym genotypowanie populacji mapującej. Jednak w związku z wybuchem pandemii COVID-19 i zmianą kosztów odczynników do sekwencjonowania, koszt usługi genotypowania techniką targetGBS znacząco wzrósł. Koszty wzrosły do takiego poziomu, że technika targetGBS przestała być konkurencyjna cenowo w stosunku do innych alternatyw dostępnych na rynku. W związku z tym firma oferująca tę usługę zdecydowała się ją całkowicie wycofać z oferty.

Po przeanalizowaniu dostępnych alternatyw zdecydowano się wykorzystać dwie metody genotypowania:

1. opartą o mikromacierze genotypowanie technologią Illumina Infinium Array w oparciu o zestaw markerów SNP Rye 5K zawierający 4800 markerów [Haseneyer i in. 2011] oraz dodatkowo wybrane markery z zestawu Rye 600k [Bauer i in. 2017] i markery w genach kandydujących. Genotypowanie wykonano usługowo przez firmę TraitGenetics, specjalizującą się w tego typu usługach. Analizy genotypowania polegały na wykonaniu wiązania próbek DNA z sondami i detekcji sygnałów hybrydyzacji. Na podstawie odczytanych wyników utworzono macierz z informacją o wariacie genetycznym badanych genotypów we wszystkich analizowanych locus, dla których udało się uzyskać sygnał hybrydyzacji.
2. opartą o sekwencjonowanie nowej generacji (NGS) wykonaną techniką DArTseq. Genotypowanie wykonano jako usługę zleconą firmie Diversity Arrays Technology Pty Ltd. Technika DArTseq, opracowana przez tę firmę, stała się jej flagową usługą oferowaną dla wielu gatunków roślin i zwierząt. Analizy polegały na wykonaniu reprezentacji genomowej poprzez trawienie genomowego DNA odpowiednią kombinacją enzymów restrykcyjnych. Do tak przygotowanych fragmentów dołączono przez ligację adaptory (fragmenty DNA zawierające informację niezbędną do hybrydyzacji fragmentów DNA z komorą przepływową oraz indeksy pozwalające na identyfikację poszczególnych próbek w późniejszej analizie wyników). Następnie wykonano wysoko przepustowe sekwencjonowanie oparte o technologię sekwencjonowania przez syntezę na urządzeniu Illumina.

Z roślin pokolenia F₂ populacji mapującej został wyizolowany materiał genetyczny (DNA). Do izolacji wykorzystano zmodyfikowaną metodę CTAB oraz kity do izolacji genomowego DNA. Po izolacji ilość oraz jakość wyizolowanego DNA sprawdzono poprzez przeprowadzenie elektroforezy a także pomiary spektrofotometryczne. Wszystkie wyizolowane próbki przeszły wewnętrzną kontrolę jakości, a mianowicie stężenie powyżej 120 ng/ul, wysoka masa cząsteczkowa oraz współczynniki absorpcji A₂₆₀/A₂₈₀ w przedziale 1.8 do 2.0 oraz A₂₆₀/A₂₃₀ > 2.0. Następnie sporządzono roztwory robocze o stężeniu wymaganym przez firmy wykonujące usługowo genotypowanie, a mianowicie 30 ng/ul dla genotypowania DArTseq oraz 120 ng/ul dla genotypowania mikromacierzami.

Analiza segregacji zidentyfikowanych loci polimorficznych w populacji F₂ żyta została przeprowadzona przy użyciu metody DArTseq. Proces ten obejmował trawienie genomowego DNA żyta enzymami restrykcyjnymi, tworząc reprezentację genomową, a następnie przygotowanie na tej podstawie bibliotek sekwencyjnych. Biblioteki te poddano wysokoprzepustowemu sekwencjonowaniu a następnie wykorzystując narzędzia bioinformatyczne wykonano identyfikację polimorfizmów SNP. Linie w populacji F₂ identyfikowano dzięki specjalnie zaprojektowanym adapterom dodawanym do oligonukleotydów przed amplifikacją. Kontrola jakości surowych odczytów oraz ich mapowanie wykonano na referencyjną sekwencję genomu żyta, wykorzystując narzędzia takie jak Bowtie2. Analizy bioinformatyczne obejmowały identyfikację wariantów SNP w poszczególnych liniach, wykorzystując specjalnie opracowany protokół analityczny. Kończącym wynikiem była macierz genotypowania dla każdego locus SNP i polimorfizmu insercji-delecji, wraz z sekwencją każdego markera i jego lokalizacją w genomie żyta [Rabanus-Wallace i in. 2021], a także parametrami statystycznymi takimi jak wartość e-value i średnia liczba odczytów.

Genotypowanie populacji F₂ żyta za pomocą techniki mikromacierzowej z zastosowaniem zestawu Rye 5K, opisanego przez Haseneyer i in. w 2011 roku, stanowiło uzupełniającą metodę do DArTseq. W tej metodzie, mikromacierze SNP (pojedynczy nukleotyd polimorfizmu) używane są do analizy genetycznej. Zestaw Rye 5K zawiera około 5000 markerów SNP, które zostały specjalnie zaprojektowane do identyfikacji wariantów genetycznych w genomie żyta. Podczas genotypowania, DNA z każdego osobnika z populacji F₂ poddawany był hybrydyzacji na mikromacierzach, co pozwoliło na jednoczesną analizę wielu markerów SNP. Dzięki temu możliwe było uzyskanie wariantu genetycznego w każdym z analizowanych loci SNP dla wszystkich analizowanych linii populacji mapującej.

Macierz z wynikami genotypowania (DArTseq oraz genotypowanie mikromacierzami) została poddana testowi Chi-kwadrat odchylenia rozkładu od wartości oczekiwanej 1:2:1. Markery spełniające kryterium istotności p-value ($\geq 0,00001$) zaimportowano do programu R i przy użyciu pakietu funkcji `mstmap.cross` z pakietu `ASMap` [Taylor i Butler 2017] poddano rozlokowaniu w grupach sprzężeń przy użyciu statystyki LOD (logarytm dziesiętny z ilorazu prawdopodobieństwa sprzężenia dwóch loci do prawdopodobieństwa, że nie są sprzężone) oraz analizy częstości rekombinacji. Rozlokowanie markerów w grupach sprzężeń zostało porównane z lokalizacją w chromosomach i kontigach loci SNP użytych do ich zaprojektowania (dla wybranych grup sprzężeń).

W wyniku genotypowania metodą DArT-seq uzyskano 8239 markerów o średniej liczbie odczytów min. 10. Jest to wynik typowy dla tego typu analiz. Genotypowanie mikromacierzami pozwoliło na uzyskanie 4851 markerów SNP. Dane o segregacji markerów dla populacji mapujących zostały zaimportowane do programu R i za pomocą funkcji `mstmap.cross` z pakietu `ASMap` utworzono grupy sprzężeń przy użyciu statystyki LOD (logarytm dziesiętny z ilorazu prawdopodobieństwa sprzężenia dwóch loci do prawdopodobieństwa, że nie są sprzężone) oraz analizy częstości rekombinacji. Metoda ta została pozwala na szybkie konstruowanie map na wiele różnych sposobów, korzystając z dostarczonych wyników genotypowania. Początkowo mapa może być zachowana w całości, lub można wybrać podzbiór wybranych grup sprzężeń, korzystając z argumentu `chr`. Ustawienie `bychr = FALSE` powoduje skupienie informacji o markerach dla wybranych grup i w razie potrzeby utworzenie nowych grup oraz optymalizację kolejności markerów w każdej z nich. Ustawienie `bychr = TRUE` zapewnia optymalne uporządkowanie markerów w każdej grupie łączenia. Powoduje to również podział grup w zależności od wartości p, podanej w wywołaniu. Algorytm pozwala na dostosowanie prognozy wartości p (`p.value`) dla klastrowania markerów do odrębnych grup (Wu et al., 2008) i jest w dużej mierze zależny od liczby osobników w populacji. Uzyskana kolejność markerów w obrębie testowanych grup sprzężeń (1R, 7R) odzwierciedlała kolejność występowania poszczególnych loci polimorficznych w sekwencji genomu żyta [Rabanus-Wallace i in 2021]. Ryciny 2-4 ilustrują rozmieszczenie zidentyfikowanych markerów SNP w genomie żyta dla odpowiednich wszystkich markerów; markerów DArT oraz markerów mikromacierzowych. Dystrybucja wszystkich markerów genetycznych (łącznie mikromacierze i DArT) pozwala zauważyć, że niektóre regiony

chromosomów mają wyższą gęstość markerów, co może wskazywać na obszary o szczególnym zainteresowaniu lub większej zmienności genetycznej.

W ramach przeprowadzonej analizy genotypowania populacji żyta z wykorzystaniem technik DArT-seq i mikromacierzy, uzyskano istotne dane dotyczące rozmieszczenia markerów genetycznych na chromosomach. Metoda DArT-seq pozwoliła na identyfikację 8239 markerów, co jest rezultatem typowym dla tego rodzaju analiz i świadczy o wysokiej przepustowości tej technologii. Z kolei technika mikromacierzy dostarczyła 4851 markerów SNP, co jest wynikiem zgodnym z oczekiwaniami dla tej technologii, oferującej większą precyzję w lokalizacji poszczególnych loci, choć przy niższej przepustowości w porównaniu do DArT-seq. Import danych o segregacji markerów do środowiska R i zastosowanie funkcji `mstmap.cross` z pakietu `ASMap` umożliwiło stworzenie map sprzężeń. Wykorzystanie statystyki LOD oraz analizy częstości rekombinacji pozwoliło na efektywne grupowanie markerów i optymalizację ich kolejności w obrębie poszczególnych grup sprzężeń. Wyniki uzyskane w badaniu są spójne z kolejnością loci polimorficznych w genomie żyta, co zostało potwierdzone w literaturze [Rabanus-Wallace i in. 2021], co świadczy o wysokiej jakości i wiarygodności genotypowania. Analiza rozmieszczenia markerów SNP w genomie żyta pokazuje zróżnicowaną dystrybucję, z wyraźnie rozróżnialnymi charakterystykami dla markerów DArT i mikromacierzowych. Równomierna dystrybucja markerów mikromacierzowych może świadczyć o stabilnym pokryciu genomu, co jest kluczowe dla celów hodowlanych i identyfikacji nowych cech fenotypowych. Z kolei markery DArT, choć rozłożone mniej równomiernie, mogą dostarczać informacji o regionach genomu mniej reprezentowanych przez mikromacierze, ujawniając obszary o potencjalnie większym zróżnicowaniu genetycznym. Obserwowane obszary o zwiększonej gęstości markerów, mogą wskazywać na regiony genomu zawierające geny kontrolujące ważne cechy agronomiczne, takie jak odporność na choroby czy adaptacja do specyficznych warunków środowiskowych. Te "wyspy" wysokiej gęstości SNP mogą być szczególnie interesujące w kontekście dalszych badań nad lokalizacją genów kandydujących [McKenna i in. 2010], ponieważ te regiony mogą zawierać znaczące informacje o dziedziczności cech. Wnioski płynące z tej analizy będą wykorzystane do dalszego rozwoju map genetycznych żyta, co jest kluczowym do osiągnięcia celów projektu w kolejnych latach i może przyczynić się do poprawy jakości i wydajności produkcji żyta. Ponadto, integracja danych z różnych platform genotypowania może prowadzić do bardziej szczegółowego zrozumienia struktury genetycznej tej rośliny, co jest zgodne z trendami w genetyce roślin [Varshney i in. 2014].

Ad 3)

Materiał do analiz stanowiło całkowite RNA wyizolowane z liści siewek wytypowanych genotypów o zróżnicowanej podatności na porażenie przez rdzę brunatną. Rośliny kielkowano w kontrolowanych warunkach w fitotronie. Inokulację przeprowadzono w 14 dniu od wysiewu. Materiał pobrano w trzech powtórzeniach biologicznych i dwóch punktach czasowych.

Całkowite RNA do analiz transkryptomu wyizolowano przy użyciu kitu Total RNA Mini (A&A Biotechnology). Ilość oraz jakość (integralność) wyizolowanego materiału sprawdzono na urządzeniu Qiaxcel Advance przy wykorzystaniu zestawu RNA QC (Qiagen). Biblioteki do sekwencjonowania przygotowano za pomocą zestawu TruSeq Stranded mRNA LT Sample Prep Kit (Illumina) według protokołu TruSeq Stranded mRNA Sample Preparation Guide, Part # 15031047 Rev. E, co pozwoliło na uzyskanie bibliotek o znanej orientacji nici (strand sensitive). Sekwencjonowanie wykonano na urządzeniu HiSeq2500 w technologii 150 PE, 60 mln odczytów. Wyniki sekwencjonowania zapisano jako pliki fastq (zawierające informację sekwencyjną oraz informację o jakości uzyskanych odczytów). Do analizy otrzymanych wyników opracowano potok analityczny w oparciu o narzędzie Nextflow (<https://www.nextflow.io/>). Zastosowanie potoku analitycznego opracowanego w oparciu o Nextflow pozwoliło na standaryzację otrzymanych wyników. Dodatkowo uzyskane wyniki były bardziej powtarzalne i wiarygodne. Pierwszą analizą potoku była kontrola jakości. Wykorzystano program fastqc. Odczyty słabej jakości $Q < 30$ oraz adaptory usunięto za pomocą programu Trimmomatic (Bolger i in. 2014). Oczyszczone odczyty za pomocą programu bowtie2 zmapowano do dwóch wersji genomu referencyjnego. Utworzone pliki sam z informacją o położeniu poszczególnych

odczytów przekształcono w posortowane pliki bam za pomocą narzędzia samtools (Li 2011). Tak przygotowane pliki bam posłużyły do przygotowania macierzy zaliczeń odczytów przypisanych do poszczególnych genów za pomocą dwóch algorytmów RSEM (transcriptome dependent) oraz kallisto (transcriptome independent). Uzyskane wyniki ekspresji znormalizowano za pomocą metody lognorm z pakietu DESeq2 (Love i in. 2019). Analizę porównawczą ekspresji genów (DE) wykonano w tym samym pakiecie programu R (Team i in. 2013). W analizach statystycznych wykorzystano model liniowy: $Y = \mu + \alpha \times \beta \times \gamma + \epsilon$. Do anotacji opracowanego transkryptomu wykorzystano: program BLAST (Altschul i in. 1990) i algorytm tblastn oraz program InterProScan. Za pomocą programu InterProScan do odtworzonych transkryptów przypisano numery GO. Na podstawie tych danych oraz wyników uzyskanych w analizach DE wykonano analizę 'enrichment analysis', która pozwoliła na wskazanie grup genów, które zmieniają swoją ekspresję pod wpływem infekcji patogenu. Obliczenia wykonano przy użyciu pakietu GOSec (Young i in. 2012) w programie R. Pakiet GOSec uwzględniał w analizie wpływ długości genu i brał poprawkę na tą wartość. W analizie GO wykorzystano test Walleniusa. Otrzymane wartości p poprawiono ze względu na wykonanie wielokrotnych porównań przy użyciu metody Benjamini-Hochberga (Benjamini i Hochberg 1995). Za istotnie nadreprezentowane uznano numery GO dla których wartość $FDR < 0.05$.

Porównano także wpływ genotypu i jego podatności na porażenie na zmiany w ekspresji genów związanych z odpowiedzią rośliny na patogen. Analizy GO wykonano na podstawie znormalizowanych danych ekspresji metodą lognorm, dodatkowo wykonano mapę PCA dla 1 i 2 składowej głównej. Na podstawie znormalizowanych danych ekspresji genów skonstruowano także mapę cieplną obrazującą współczynniki korelacji Pearsona pomiędzy poszczególnymi próbkami.

Dane z sekwencjonowania wykorzystano także w analizach VariantCalling w celu identyfikacji polimorfizmów pojedynczego nukleotydu, a także krótkich insercji i delecji (INDEL). Analizę VariantCalling wykonano za pomocą potoku analitycznego GATK. Otrzymane odczyty porównano do otrzymanego de novo transkryptomu za pomocą programu bowtie2. Uzyskane w wyniku tych analiz pliki bam wykorzystano do poszukiwania mutacji punktowych za pomocą programu HaplotypeCaller. Otrzymane surowe SNP następnie przefiltrowano na podstawie uzyskanych parametrów mapowania i tylko najlepsze SNP wykorzystano w dalszych pracach. Analizy bioinformatyczne i obliczenia statystyczne wykonano na serwerze „bioscienta” należącym do Katedry Genetyki Hodowli Roślin i Nasiennictwa Uniwersytetu Przyrodniczego we Wrocław, wyposażonym w odpowiednie oprogramowanie bioinformatyczne i statystyczne.

W badaniu z powodzeniem wyodrębniono RNA z ośmiu próbek (UP1 do UP8), osiągając koncentrację od 6 do 32 g. Jakość wyizolowanego RNA była wysoka, co potwierdzają współczynniki RIN przekraczające 8. Efektywność wybranej metody izolacji RNA dla badanego materiału została potwierdzona, co było kluczowe dla dalszego sekwencjonowania transkryptomu. Dzięki uzyskaniu RNA o wysokiej jakości, proces przygotowania bibliotek przebiegał sprawnie, co zaowocowało ich wysoką jakością i pomyślnym sekwencjonowaniem. Sekwencjonowanie każdej z bibliotek dostarczyło od 43 do 44 milionów odczytów oraz od 7,2 do 7,8 Gbp odczytanych nukleotydów. Około 95% odczytów charakteryzowało się jakością Phred powyżej 30, świadcząc o sukcesie sekwencjonowania, podsumowano w tabeli 2. Wysoka jakość odczytów umożliwiła przeprowadzenie analiz różnic ekspresji genów (DE). Mapowanie wyników pozwoliło na przypisanie większości odczytów do konkretnych lokalizacji w genomie referencyjnym, z co najmniej 80% odczytów zmapowanych do jednego miejsca dla każdej próbki, co umożliwiło dalsze porównawcze analizy ekspresji. Po załadowaniu danych do R i użyciu pakietu DESeq2 obliczono współczynniki normalizacji dla bibliotek (Tabela 4), które wskazywały na podobną głębokość sekwencjonowania we wszystkich próbkach. Spójność wyników między powtórzeniami biologicznymi, zilustrowana na rysunku 1, potwierdza poprawność przeprowadzonych eksperymentów. Analiza porównawcza ekspresji ujawniła istotną zmianę ekspresji w 3492 genach ($FDR < 0,01$) o przynajmniej dwukrotnie wyższej lub niższej ekspresji w roślinach uzyskanych klasycznymi metodami.

W oparciu o przedstawione wyniki, nasza dyskusja koncentruje się na kluczowych aspektach biologii molekularnej i genetyki, ze szczególnym uwzględnieniem izolacji RNA, sekwencjonowania transkryptomu i analizy różnic ekspresji genów (DE). Wykazaliśmy, że zastosowana metoda izolacji RNA była skuteczna, co potwierdzają wysokie współczynniki RIN (powyżej 8) oraz znaczące koncentracje RNA (6 do 32 g). Jakość RNA jest kluczowa dla sukcesu sekwencjonowania transkryptomu, co znajduje potwierdzenie w literaturze [Schroeder i in., 2006; Fleige i Pfaffl, 2006]. Wyniki te podkreślają znaczenie zastosowania odpowiednich metod izolacji RNA w celu zapewnienia wiarygodności wyników molekularnych analiz. Uzyskano od 43 do 44 milionów odczytów dla każdej z bibliotek, z około 95% odczytów charakteryzujących się jakością Phred powyżej 30. Ta ilość i jakość danych sekwencjonowania odpowiada aktualnym standardom i wskazuje na wysoką efektywność procesu (Metzker, 2010). Takie wyniki świadczą o skuteczności naszej metodyki i jej przydatności w analizie transkryptomowej. Stosowanie pakietu DESeq2 do analizy DE i identyfikacji istotnych zmian ekspresji w 3492 genach jest zgodne z aktualnymi standardami (Love et al., 2014). Wyniki te potwierdzają skuteczność metody w identyfikacji kluczowych genów różnicujących badane grupy, co ma znaczenie dla zrozumienia mechanizmów molekularnych leżących u podstaw obserwowanych fenotypów. Analizy dotyczące genów związanych z odpornością, takich jak geny klasy TIR-NBS-LRR czy geny transporterów ABC, są zgodne z wcześniejszymi badaniami nad mechanizmami obronnymi roślin (Jones i Dangl, 2006; Kangasjärvi et al., 2012). Wskazuje to na potencjalną rolę tych genów w adaptacji roślin do stresów biotycznych, co otwiera nowe możliwości dla dalszych badań w tej dziedzinie. Podsumowując, nasze badanie skutecznie wykorzystuje zaawansowane techniki biologii molekularnej do uzyskania dogłębnego wglądu w zmiany ekspresji genów. Osiągnięta wysoka jakość danych RNA-seq, połączona z kompleksową analizą statystyczną, umożliwia głębokie zrozumienie mechanizmów molekularnych. Wyniki te mają istotne implikacje dla dalszych badań związanych z analizą mechanizmu molekularnej odporności żyta na rdzę brunatną.

Literatura:

- Bujak, H., Nowosad, K., & Łącka, A. (2019). Poszukiwanie źródeł genetycznej odporności na mączniaka i rdzę w kolekcji linii, rodów i odmian żyta. INSTYTUTU HODOWLI I AKLIMATYZACJI ROŚLIN NR 295/2021, 181.
- Pietrusińska, A, Żurek, M, Piechota, U, Słowacki, P, & Smolińska, K (2018). Searching for diseases resistance sources in old cultivars, landraces and wild relatives of cereals. A review. *Agronomy Science*, 73(4).
- Sawinska, Z. (2011). Influence of powdery mildew and brown rust on winter rye yielding. *Progress in Plant Protection*, 51(3), 1193-1197.
- Weber, R., Bujak, H., Nowosad, K., Gacek, E., & Kotowicz, L. (2015). Analiza zmienności porażenia odmian żyta ozimego przez grzyb *Puccinia recondita* na Dolnym Śląsku. *Polish Journal of Agronomy*, 23, 82-87.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., ... & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, 20(9), 1297-1303.
- Li, Y., Haseneyer, G., Schön, C. C., Ankerst, D., Korzun, V., Wilde, P., & Bauer, E. (2011). High levels of nucleotide diversity and fast decline of linkage disequilibrium in rye (*Secale cereale* L.) genes involved in frost response. *BMC Plant Biology*, 11, 1-14.
- Rabanus-Wallace, M. T., Hackauf, B., Mascher, M., Lux, T., Wicker, T., Gundlach, H., ... & Stein, N. (2021). Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. *Nature Genetics*, 53(4), 564-573.
- Taylor, J., & Butler, D. (2017). R package ASMap: efficient genetic linkage map construction and diagnosis. arXiv preprint arXiv:1705.06916.

- Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., ... & Cook, D. R. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nature biotechnology*, 31(3), 240-246.
- Wu, Y., Bhat, P. R., Close, T. J., & Lonardi, S. (2008). Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS genetics*, 4(10), e1000212.
- Altschul, Stephen F, Warren Gish, Webb Miller, Eugene W Myers, and David J Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3). Elsevier: 403–10.
- Baldwin, S., Pither-Joyce, M., Wright, K., Chen, L., & McCallum, J. (2012). Development of robust genomic simple sequence repeat markers for estimation of genetic diversity within and among bulb onion (*Allium cepa* L.) populations. *Molecular Breeding*, 30(3), 1401-1411.
- Bolger, Anthony M, Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30 (15). Oxford University Press: 2114–20.
- Conesa, Ana, Pedro Madrigal, Sonia Tarazona, David Gomez-Cabrero, Alejandra Cervera, Andrew McPherson, Michał Wojciech Szczęśniak, et al. 2016. "A Survey of Best Practices for Rna-Seq Data Analysis." *Genome Biology* 17 (1). BioMed Central: 13.
- Fleige, S., & Pfaffl, M. W. (2006). RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular aspects of medicine*, 27(2-3), 126-139.
- Jones, J. D., & Dangl, J. L. (2006). The plant immune system. *nature*, 444(7117),323-329.
- Kangasjärvi, S., Neukermans, J., Li, S., Aro, E. M., & Noctor, G. (2012). Photosynthesis, photorespiration, and light signalling in defence responses. *Journal of experimental botany*, 63(4), 1619-1636.
- Kim, Daehwan, Joseph M Paggi, Chanhee Park, Christopher Bennett, and Steven L Salzberg. 2019. "Graph-Based Genome Alignment and Genotyping with Hisat2 and Hisat-Genotype." *Nature Biotechnology* 37 (8). Nature Publishing Group: 907–15.
- Kohorn, B. D., & Kohorn, S. L. (2012). The cell wall-associated kinases, WAKs, as pectin receptors. *Frontiers in plant science*, 3, 88.
- Laluk, K., AbuQamar, S., & Mengiste, T. (2011). The Arabidopsis mitochondria-localized pentatricopeptide repeat protein PGN functions in defense against necrotrophic fungi and abiotic stress tolerance. *Plant Physiology*, 156(4), 2053-2068.
- Li, Heng. 2011. "A Statistical Framework for Snp Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation from Sequencing Data." *Bioinformatics* 27 (21). Oxford University Press: 2987–93.
- Love, Michael I, Simon Anders, and Wolfgang Huber. 2019. "Analyzing Rna-Seq Data with Deseq2."
- Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature reviews genetics*, 11(1), 31-46.
- Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., ... & Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC molecular biology*, 7(1), 1-14.

Shen, Y., Liu, N., Li, C., Wang, X., Xu, X., Chen, W., ... & Zheng, W. (2017). The early response during the interaction of fungal phytopathogen and host plant. *Open biology*, 7(5), 170057.

Team, R Core, and others. 2013. "R: A Language and Environment for Statistical Computing." Vienna, Austria.

Young, Matthew D, Matthew J Wakefield, Gordon K Smyth, and Alicia Oshlack. 2012. "Goseq: Gene Ontology Testing for Rna-Seq Datasets." R Bioconductor.